

Methodology article

Open Access

Quantitative trait locus analysis of hybrid pedigrees: variance-components model, inbreeding parameter, and power

Gulnara R Svischeva

Address: Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science, Lavrentiev Ave, 10, Novosibirsk, 630090, Russia

Email: Gulnara R Svischeva - gulsvi@mail.ru

Published: 26 July 2007

Received: 30 January 2007

BMC Genetics 2007, 8:50 doi:10.1186/1471-2156-8-50

Accepted: 26 July 2007

This article is available from: <http://www.biomedcentral.com/1471-2156/8/50>

© 2007 Svischeva; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: For the last years reliable mapping of quantitative trait loci (QTLs) has become feasible through linkage analysis based on the variance-components method. There are now many approaches to the QTL analysis of various types of crosses within one population (breed) as well as crosses between divergent populations (breeds). However, to analyse a complex pedigree with dominance and inbreeding, when the pedigree's founders have an inter-population (hybrid) origin, it is necessary to develop a high-powered method taking into account these features of the pedigree.

Results: We offer a universal approach to QTL analysis of complex pedigrees descended from crosses between outbred parental lines with different QTL allele frequencies. This approach improves the established variance-components method due to the consideration of the genetic effect conditioned by inter-population origin and inbreeding of individuals. To estimate model parameters, namely additive and dominant effects, and the allelic frequencies of the QTL analysed, and also to define the QTL positions on a chromosome with respect to genotyped markers, we used the maximum-likelihood method. To detect linkage between the QTL and the markers we propose statistics with a non-central χ^2 -distribution that provides the possibility to deduce analytical expressions for the power of the method and therefore, to estimate the pedigree's size required for 80% power. The method works for arbitrarily structured pedigrees with dominance and inbreeding.

Conclusion: Our method uses the phenotypic values and the marker information for each individual of the pedigree under observation as initial data and can be valuable for fine mapping purposes. The power of the method is increased if the QTL effects conditioned by inter-population origin and inbreeding are enhanced. Several improvements can be developed to take into account fixed factors affecting trait formation, such as age and sex.

Background

The wide application of DNA markers scattered along the genome together with the rapid development of statistical methods provides reliable localization of quantitative trait loci (QTLs). There are now many approaches to QTL

analysis of various types of crosses within one population (breed) as well as crosses between divergent populations (breeds) [1-5]. One of the most powerful approaches to QTL mapping is the variance-components method. In this method, variability among trait observations from indi-

viduals within pedigrees is expressed in terms of the effect caused by an unobservable trait-affecting major locus, of the polygenic effect, and of the residual non-genetic effect [2,3,6-13]. The effect attributable to a locus linked to a marker is a function of the additive and dominance components of variance of the locus, the recombination fraction, and the portion of alleles that are identical by descent (IBD) at the marker for each pair of individuals. The polygenic variance component depends only on the relationship between the relative pair.

If the pedigree analysed comes from a population with an identical distribution of genotypes for all the members of the pedigree and with an identical environmental influence on phenotypes, then the covariance between trait values of related pairs is the weighted sum of the variance components identical for all individuals [10,14-16]. The presence of the marker information makes it possible to separate the variance component caused by the locus linked to the marker from the polygenic variance component, and to test the significance of major locus contribution with respect to trait polymorphism.

Crosses between individuals from divergent populations (breeds) that differ by trait distribution are often used in investigations of traits of livestock breeding, laboratory and domestic animals, and studies of human hereditary diseases. A set of statistical methods for QTL mapping was developed in which initial materials that are backcrosses or the F_2 -generation descended from inbred lines were used [17-21]. Recently several studies devoted to the analysis of crosses between outbreed lines has been reported [2,3]. One of these statistical methods, known as the segment mapping method [2], is based on division of the genome of hybrid individuals into segments. Here, genetic covariance of a trait is defined for each segment and depends on the variance of initial breeds and the percentage of genetic material of these breeds in this segment. However, this method does not take into consideration such effects as domination and inbreeding. On the contrary, another method developed in [3] assumes the presence of these effects and allows us to find distinctions in genotype frequencies of the major locus analysed between the crossed breeds. The essence of this method is that the genetic covariance of any two individuals is expressed as a non-linear function of the probability of up to 15 possible identity modes differing by the allele origin of the locus. The disadvantage of this method is its inherent complicated calculations.

The objective of the present study is to present another high-powered theoretical approach to analyse data from crosses between outbreed lines using marker information. This approach is based on the variance-components method that takes into account dominance and inbreed-

ing and uses all the pedigree information available. This study is structured as follows. First, we formulate assumptions about the genetic inter-population nature of the trait so that these assumptions allow us to prove the genetic model chosen and the distribution of phenotypes in the pedigree. Second, we develop a universal way for decomposing variance and covariance into equi-type components, so that weighting factors at these components depend on the degree of relationship and the recombination frequency between the marker and the locus, and can be obtained from joint distribution of IBD-alleles of the QTL and the marker [7]. This ensures that we derive the exact analytical expressions of variance components for different types of relative pairs. Third, we obtain analytical expressions for the power of our method without simulation data. The method is demonstrated by an example of hybrid sibships, which are widely popular in experimental designs.

Results

The genetic model

A general explanatory multi-locus model describing the quantitative trait for the i th individual of a hybrid pedigree is

$$X_i = \mu_i + g_i + G_i + e_i,$$

where μ is the overall mean, g and G denote independent effects conditioned by the influence of QTLs (major locus and polygene, respectively), and e denotes the environmental effect. However, since the contribution of the major locus to the trait studied has no priorities in relation to other loci listed in the polygene, we will consider a simplified mono-locus model, which could be easily extended to general cases without major difficulties.

For the analysis of crosses between two divergent populations, P_1 and P_2 , it is necessary to consider additional assumptions about equi-type distribution of the trait in the parental populations, P_1 and P_2 , and in the hybrid pedigree, $P_1 \times P_2$. We assume that QTL contributions to trait formation do not depend on the population origin of the individuals, and that crossed initial populations differ by unequal QTL allele frequencies, p_1 for P_1 and p_2 for P_2 [3,23]. In addition, we assume that the Hardy-Weinberg equilibrium is carried out for the P_1 and P_2 populations.

We consider the QTL analysed to be an autosomal locus with alleles A and B , and its genotypic values g , equal to d , a and $-a$ for heterozygotic and homozygotic individuals of alternative forms, respectively. AA_i (AB_i or BB_i) denotes that individual i has genotype AA (AB , or BB). The distribution of frequencies of the QTL genotypes, AA_i , AB_i and BB_i , for the i th inbred hybrid individual can be expressed

by allelic frequencies of his (her) father ($p(A_f)$, $p(B_f)$) and mother ($p(A_m)$, $p(B_m)$):

$$\begin{aligned} p(AA_i) &= p(A_f)p(A_m) + \tau, \\ p(AB_i) &= p(A_f)p(B_m) + p(B_f)p(A_m) - 2\tau, \\ p(BB_i) &= p(B_f)p(B_m) + \tau, \end{aligned} \quad (1)$$

where τ denotes a positive-definite parameter of inbreeding caused by the correlation between the uniting gametes of the inbred individual, and shows the difference in genotype frequencies between non-inbred and inbred homozygous individuals [24]. If the individual is non-inbred, then $\tau = 0$. We have considered various types of inbred crosses differing from each other in structure of inbred loops and represented the derivations of the parameter of inbreeding (see section "Parameter of inbreeding" and Appendix). The allelic frequencies, $p(A_i)$ and $p(B_i)$, are determined through the distribution of genotypes as:

$$\begin{aligned} p(A_i) &= p(AA_i) + \frac{1}{2}p(AB_i), \\ p(B_i) &= p(BB_i) + \frac{1}{2}p(AB_i). \end{aligned} \quad (2)$$

Furthermore, the allelic frequencies of the i th hybrid individual can be expressed in terms of the allelic frequencies of the initial populations, p_1 and p_2 , and a parameter, ε_{i1} ($\varepsilon_{i2} = 1 - \varepsilon_{i1}$), called portion of "blood" of the population P_1 (P_2) [23]:

$$p(A_i) = \varepsilon_{i1} p_1 + (1 - \varepsilon_{i1}) p_2.$$

We admit that the trait values of individuals from a hybrid pedigree, as well as from P_1 and P_2 populations, have a multi-normal distribution that is parameterised by an expectation vector and a covariance matrix [25]. If the influence of the environment is identical for all hybrid individuals, then without sacrificing the model generality we can assume that environmental effects for all individuals are random effects distributed by the normal law with identical parameters of distribution, $N(0, \text{Var}_e)$ [10].

Parameter of inbreeding

We have defined the parameter of inbreeding for two particular pedigrees with different inbred loops, and general-

ized the conclusions drawn to all other pedigrees. Let the first pedigree under review include the shortest inbred loop with a single common ancestor for parents of the inbred individual (Figure 1a). In this case, the inbred individual descends from a cross of the related pair of "parent-offspring". To find τ , we have considered a similar pedigree with the same structure but without inbreeding (Figure 1b). We need to determine the distributions of genotype frequencies for the 4th individuals from inbred and outbred crosses through allelic frequencies of the pedigree's founders, $p(A_1)$ and $p(A_2)$, and to compare them with each other. For this purpose, we first defined the distribution of genotype frequencies for the 3rd individuals, which are non-inbred for both pedigrees (using formulas (1) at $\tau = 0$):

$$\begin{aligned} p(AA_3) &= p(A_1)p(A_2), \\ p(AB_3) &= p(A_1)p(B_2) + p(B_1)p(A_2), \\ p(BB_3) &= p(B_1)p(B_2). \end{aligned}$$

As a consequence, allelic frequencies for the 3rd individuals are equal to $p(A_3) = 1/2 (p(A_1) + p(A_2))$ and $p(B_3) = 1/2 (p(B_1) + p(B_2))$ in accordance with formulas (2). Therefore we can write the distribution of genotype frequencies for the 4th individual from the outbred cross (Figure 1b):

$$\begin{aligned} p(AA_4) &= \frac{1}{2}(p(A_1) + p(A_2))p(A_2), \\ p(AB_4) &= \frac{1}{2}(p(A_1) + p(A_2))p(B_2) + \frac{1}{2}(p(B_1) + p(B_2))p(A_2), \\ p(BB_4) &= \frac{1}{2}(p(B_1) + p(B_2))p(B_2). \end{aligned}$$

We deal with the conditional distribution of genotype frequencies upon analysing the 4th individual from the inbred cross. Let us therefore fix the genotype of the 2nd individual, $g_2 = AA, AB$ or BB with a probability $p(AA_2)$, $p(AB_2)$ or $p(BB_2)$. The conditional genotype probabilities for the 3rd and 4th individuals are thus easily calculated and are presented in Tables 1 and 2. According to the formula of total probabilities, the unconditional probability of the genotypes for the 4th inbred individual, $p^{\text{inb}}(g_4)$, is equal to:

$$\begin{aligned} p^{\text{inb}}(g_4) &= \sum_{g_2} p(g_2) \sum_{g_3} p(g_4 / g_3, g_2) p(g_3 / g_2) \\ &= \sum_{g_2} p(g_4 / g_2) p(g_2). \end{aligned}$$

To estimate the parameter of inbreeding, we compared genotype probabilities of the 4th individuals from inbred and outbred crosses, $p^{\text{inb}}(AA_4)$ and $p(AA_4)$:

$$\begin{aligned} \tau &= p^{\text{inb}}(AA_4) - p(AA_4) \\ &= \frac{1}{4}(p(A_2)p(BB_2) + p(B_2)p(AA_2)). \end{aligned}$$

Table 1: Conditional distribution of genotype frequencies of the QTL for an individual under the given genotype of his (her) parent

$p(g_3/g_2)$	$g_2 = AA$	$g_2 = AB$	$g_2 = BB$
$p(AA_3/g_2)$	$p(A_1)$	$1/2 p(A_1)$	0
$p(AB_3/g_2)$	$p(B_1)$	$1/2$	$p(A_1)$
$p(BB_3/g_2)$	0	$1/2 p(B_1)$	$p(B_1)$

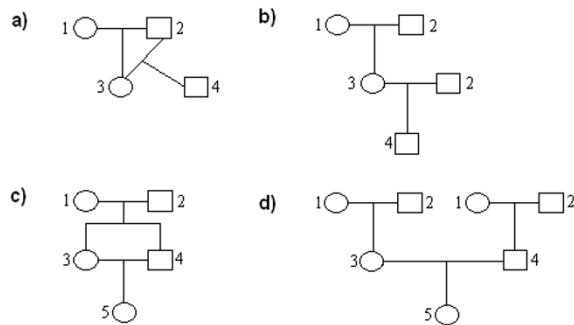


Figure 1
Pedigrees including the shortest inbred loop with a single and two common ancestors and pedigrees without the loop. a, b) Pedigrees with and without the loop formed as a result of inbred cross between parent and offspring, respectively. c, d) Pedigrees with and without the loop formed as a result of inbred cross between sibs, respectively.

The parameter of inbreeding may be generalized to any type of crosses with a single common ancestor:

$$\tau = (1/2)^k f(g_o), \quad (3)$$

where $(1/2)^k$ is the degree of relationship of the parents of the inbred individual, and $f(g_o)$ is the function of genotype frequencies of their common ancestor, g_o :

$$f(g_o) = (p(A_o) p(BB_o) + p(B_o) p(AA_o))/2. \quad (4)$$

The proof of the validity of formulas (3–4) is presented in the Appendix.

Let the second pedigree examined include the shortest inbred loop with two common ancestors of parents of an

inbred individual (Figure 1c). In this case, an inbred individual descends from a cross of sibs. To find τ , we considered a similar pedigree with the same structure but without inbreeding (Figure 1d). We need to determine the distributions of the genotype frequencies of the 5th individuals from inbred and outbred crosses through allelic frequencies of the pedigree's founders, $p(A_1)$ and $p(A_2)$, and to compare them with each other.

At the outbred cross (Figure 1d), genotype frequencies of the 3rd and 4th individuals, $p(g_3)$ and $p(g_4)$, are identical, and are calculated by formulas (1), when $\tau = 0$. In this case, the 3rd and 4th individuals transfer an allele A (B) to the offspring with equal probabilities, $p(A_3) = p(A_4) = 1/2 (p(A_1) + p(A_2))$ ($p(B_3) = p(B_4) = 1/2 (p(B_1) + p(B_2))$). The distribution of genotype frequencies for the 5th individual is thus:

$$\begin{aligned} p(AA_5) &= \frac{1}{4} (p(A_1) + p(A_2))^2, \\ p(AB_5) &= \frac{1}{2} (p(A_1) + p(A_2)) (p(B_1) + p(B_2)), \\ p(BB_5) &= \frac{1}{4} (p(B_1) + p(B_2))^2. \end{aligned}$$

At the inbred cross (Figure 1c), the 3rd and 4th individuals as well as their inbred offspring (the 5th individual) therefore have conditional genotype frequencies under fixed genotypes of the 1st and 2nd individuals with the probabilities $p(AA_1)$, $p(AB_1)$ or $p(BB_1)$, and $p(AA_2)$, $p(AB_2)$ or $p(BB_2)$ (Tables 3 and 4). Knowing the conditional genotype frequencies for the 5th individual, we can define unconditional genotype frequencies using the formula of total probabilities:

$$p^{\text{inb}}(g_5) = \sum_{g_1, g_2} p(g_5 / g_1, g_2) p(g_1) p(g_2).$$

The parameter of inbreeding is then equal to:

$$\begin{aligned} \tau &= p^{\text{inb}}(AA_5) - p(AA_5) \\ &= \frac{1}{8} (p(A_1) p(BB_1) + p(B_1) p(AA_1) + p(A_2) p(BB_2) + p(B_2) p(AA_2)), \end{aligned}$$

and can be generalized to all inbred individuals having parents with two common ancestors as:

Table 2: Conditional distribution of genotype frequencies of the QTL for an inbred individual originated from cross "parent-offspring" under the given genotype of his (her) parent

$p(g_4/g_2)$	$g_2 = AA$	$g_2 = AB$	$g_2 = BB$	$p^{\text{inb}}(g_4)$
$p(AA_4/g_2)$	$1/2 (p(A_1) + 1)$	$1/4 (p(A_1) + 1/2)$	0	$1/2 [p(A_1) p(A_2) + p(A_2) - 1/4 p(AB_2)]$
$p(AB_4/g_2)$	$1/2 p(B_1)$	$1/2$	$1/2 p(A_1)$	$1/2 [p(B_1) p(A_2) + p(A_1) p(B_2) + 1/2 p(AB_2)]$
$p(BB_4/g_2)$	0	$1/4 (p(B_1) + 1/2)$	$1/2 (p(B_1) + 1)$	$1/2 [p(B_1) p(B_2) + p(B_2) - 1/4 p(AB_2)]$

Table 3: Conditional distribution of genotype frequencies of the QTL for sibs under the given genotypes of their parents

$p(g_{3,4}/g_1, g_2)$	$g_1 \times g_2$					
	AA \times AA	BB \times BB	AB \times AB	BB \times AA, AA \times BB	AB \times AA, AA \times AB	BB \times AB, AB \times BB
$p(AA_{3,4}/g_1, g_2)$	1	0	1/4	0	1/2	0
$p(AB_{3,4}/g_1, g_2)$	0	0	1/2	1	1/2	1/2
$p(BB_{3,4}/g_1, g_2)$	0	1	1/4	0	0	1/2

$$\tau = (1/2)^k (f(g_{10}) + f(g_{20}))/2, \quad (5)$$

where $(1/2)^k$ is the degree of relationship of the parents of the inbred individual, and $f(g_{io})$ for $i = 1, 2$ is the function of genotype frequencies, g_{io} , of their i th common ancestor calculated by formula (4). The proof of the validity of formula (5) is presented in the Appendix.

Partitioning genetic covariance into components

For the QTL effect, we deduce the formulas of expectation, $E_{QTL} = a p(AA_i) + d p(AB_i) - a p(BB_i)$, and variance, $Var_{QTL} = a^2 p(AA_i) + d^2 p(AB_i) + a^2 p(BB_i) - E_{QTL}^2$, depending on the set of parameters $\{a, d, p(A_f), p(A_m), \text{ and } \tau\}$:

$$E_{QTL} = \frac{1}{2}(a+d)(1-2p(B_f)p(B_m)) - \frac{1}{2}(a-d)(1-2p(A_f)p(A_m)) - 2d\tau,$$

$$Var_{QTL} = p(A_f)p(B_f)((a+d)p(B_m) + (a-d)p(A_m))^2$$

$$+ p(A_m)p(B_m)((a-d)p(A_f) + (a+d)p(B_f))^2 + 4p(A_f)p(B_f)p(A_m)p(B_m)d^2$$

$$+ 2\tau \frac{1}{4}((a-d)(p(A_f) + p(A_m)) + (a+d)(p(B_f) + p(B_m)))^2 + d^2(p(A_f) - p(A_m))^2 - 4d^2\tau^2.$$

To partition the genetic variance into components, we modernized the approach of Amos and Elston [7] adapting it for hybrid pedigree analyses.

Let $Y_j = (X_{1j} - X_{2j})^2$, where X_{1j} and X_{2j} are phenotypic values of a trait measured on individuals 1 and 2 of the j th related pair. We denote the proportion (0, 1/2 or 1) of the alleles identical by descent at the QTL and the marker as π_{QTLj} and π_{Mj} , respectively. Let \hat{h}_{Mkj} be the estimated probability that the j th pair of individuals shares the k IBD-alleles ($k = 0, 1$ or 2) at the marker. Then the π_{Mj} can be estimated as $\hat{\pi}_{Mj} = \hat{h}_{M2j} + \frac{1}{2}\hat{h}_{M1j}$. Amos and Elston [7] have shown that covariance, Cov, can be obtained by:

$$Cov(X_{1j}, X_{2j} | \hat{\pi}_{Mj}) = \frac{1}{2}[Var(X_{1j}) + Var(X_{2j})]$$

$$- \frac{1}{2} \sum_{\pi_{QTLj}} [E(Y_j | \pi_{QTLj}) \sum_{\pi_{Mj}} Pr(\pi_{QTLj} | \pi_{Mj}) Pr(\pi_{Mj} | \hat{\pi}_{Mj})],$$

where the conditional probabilities $Pr(\pi_{Mj} | \hat{\pi}_{Mj})$ are calculated from the information given on the marker genotypes, whereas the conditional probabilities $Pr(\pi_{QTLj} | \pi_{Mj})$ depend on the genetic relationship of the relative pair and the recombination frequency between the QTL and the

marker, and have already been specified for many types of relatives [7,26]. To determine $E(Y_j | \pi_{QTLj})$, we list all 9 possible values of Y_j at various QTL genotypes and define the probabilities $Pr(Y_j | \pi_{QTLj})$ for $\pi_{QTLj} = 0, 1/2$ or 1 :

$$E(Y_j | \pi_{QTLj}) = \sum_{i=1}^9 [Y_i Pr(Y_i | \pi_{QTLj})]. \quad (6)$$

Let us review in detail the process of finding genetic covariance and its components using the example of sib-pair as the most often used in studies of hereditary diseases.

The analysis of sib-pair

The sibs present a special interest for researchers, since they are well known to belong to two-linear relatives. We deduced the formulas for the conditional probabilities $Pr(Y_j | \pi_{QTLj})$ at $\pi_{QTLj} = 0, 1/2$ or 1 (Table 5), to derive the expressions of expectations $E(Y_j | \pi_{QTLj})$ by formula (6):

$$E(Y_j | \pi_{QTLj} = 1) = \xi_j^2,$$

$$E(Y_j | \pi_{QTLj} = 1/2) = \frac{1}{2}(a-d)^2(p(A_f)p(AB_m) + p(AB_f)p(A_m))$$

$$+ \frac{1}{2}(a+d)^2(p(B_f)p(AB_m) + p(AB_f)p(B_m)) + \xi_j^2,$$

$$E(Y_j | \pi_{QTLj} = 0) = (a-d)^2p(A_f)p(AB_m) + p(AB_f)p(A_m)$$

$$+ (a+d)^2(p(B_f)p(AB_m) + p(AB_f)p(B_m)) - 2d^2p(AB_f)p(AB_m) + \xi_j^2, \quad (7)$$

where ξ_j is the trait value difference caused by the environment. Apparently, a dependence on the parameter of inbreeding is not present in formulas (7) for $E(Y_j | \pi_{QTLj})$. This means that the components caused by inbreeding are identical for covariance and variance.

Haseman and Elston [26] have shown that

$$E(Y_j | \pi_{QTLj} = 1) = Z,$$

$$E(Y_j | \pi_{QTLj} = 1/2) = A + 2D + Z, \quad (8)$$

$$E(Y_j | \pi_{QTLj} = 0) = 2A + 2D + Z,$$

where Z is the environmental variance component; A and D are additive and dominance variance components, respectively. When the pedigree under consideration is

Table 4: Conditional distribution of genotype frequencies of the QTL for an inbred individual originated from the cross of sibs under the given genotypes of parents of sibs

$p(g_5/g_1, g_2)$	$g_1 \times g_2$						$p^{inb}(g_5)$
	AA × AA	BB × BB	AB × AB	BB × AA, AA × BB	AB × AA, AA × AB	BB × AB, AB × BB	
$p(AA_5/g_1, g_2)$	1	0	1/4	1/4	9/16	1/16	$1/4[p(A_1) + p(A_2) - 1/4p(AB_1) - 1/4p(AB_2) + 2p(A_1)p(A_2)]$
$p(AB_5/g_1, g_2)$	0	0	1/2	1/2	3/8	3/8	$1/2[p(B_1)p(A_2) + p(A_1)p(B_2) + 1/4p(AB_1) + 1/4p(AB_2)]$
$p(BB_5/g_1, g_2)$	0	1	1/4	1/4	1/16	9/16	$1/4[p(B_1) + p(B_2) - 1/4p(AB_1) - 1/4p(AB_2) + 2p(B_1)p(B_2)]$

pure ($p = p(A_f) = p(A_m)$ and $q = 1 - p$), A and D are defined as:

$$A = 2 p q [(a - d)p + (a + d)q]^2,$$

$$D = 4 p^2 q^2 d^2.$$

Let us find similar A and D parameters for hybrid sibs. In this case, to determine the covariance between sibs, weighting factors at components A and D are standard functions that depend on the proportion of marker IBD-alleles between sibs and the recombination frequency [10,26]:

$$f(\pi_{Mj}, \theta) = 1 - \Psi - (1 - 2\Psi)\pi_{Mj},$$

$$g(\pi_{Mj}, \theta) = (1 - \Psi)^2 - 2(1 - \Psi)(1 - 2\Psi)\pi_{Mj} + (1 - 2\Psi)^2 h_{M2j},$$

where $\Psi = \theta^2 + (1 - \theta)^2$.

By equating expressions (7) and (8) we deduce formulas of A and D for hybrid sibs:

$$A = p(AB_f)[\frac{1}{2}((a - d)p(A_m) + (a + d)p(B_m))^2 - d^2\beta_m]$$

$$+ p(AB_m)[\frac{1}{2}((a - d)p(A_f) + (a + d)p(B_f))^2 - d^2\beta_f],$$

$$D = p(AB_f)p(AB_m)d^2,$$

$$Z = \xi_j^2,$$

where $\beta_p = (p(AB_p) - 2p(A_p)p(B_p))$ at $p = m, f$.

The trait variance can be partitioned into components:

$$\text{Var} = A + D + R + Z,$$

where R is the residual component caused by inter-population origin and inbreeding of sibs): (R = 0 for "pure" non-inbred sibs):

$$R = d^2\beta_f\beta_m - \frac{1}{2}\beta_m((a - d)p(A_f) + (a + d)p(B_f))^2 - \frac{1}{2}\beta_f((a - d)p(A_m) + (a + d)p(B_m))^2 + 2\frac{1}{4}((a - d)(p(A_f) + p(A_m)) + (a + d)(p(B_f) + p(B_m)))^2 + d^2(p(A_f) - p(A_m))^2 - 4d^2\tau^2.$$

Table 5: Conditional probability distribution of Y_j values for pair of sibs

Genotypes of sib-pair	Y_j	Conditional probability $\Pr(Y_j \pi_{QTij})$		
		$\pi_{QTij} = 0$	$\pi_{QTij} = 1/2$	$\pi_{QTij} = 1$
AA-AA	ξ_j^2	$p(AA_i)p(AA_m) + \tau$	$1/2(p(AA_i)p(A_m) + p(A_i)p(AA_m)) + \tau$	$p(A_i)p(A_m) + \tau$
BB-BB	ξ_j^2	$p(BB_i)p(BB_m) + \tau$	$1/2(p(BB_i)p(B_m) + p(B_i)p(BB_m)) + \tau$	$p(B_i)p(B_m) + \tau$
AB-AB	ξ_j^2	$p(AA_i)p(BB_m) + p(BB_i)p(AA_m) + 1/2p(AB_i)p(AB_m) - 2\tau$	$p(B_i)p(A_m) + p(A_i)p(B_m) - 1/4[p(AB_i) + p(AB_m)] - 2\tau$	$p(A_i)p(B_m) + p(B_i)p(A_m) - 2\tau$
AA-AB	$(a - d + \xi_j)^2$	$1/2(p(AB_i)p(AA_m) + p(AA_i)p(AB_m))$	$1/4(p(A_i)p(AB_m) + p(AB_i)p(A_m))$	0
AB-AA	$(-a + d + \xi_j)^2$	$1/2(p(AB_i)p(AA_m) + p(AA_i)p(AB_m))$	$1/4(p(A_i)p(AB_m) + p(AB_i)p(A_m))$	0
AB-BB	$(a + d + \xi_j)^2$	$1/2(p(AB_i)p(BB_m) + p(BB_i)p(AB_m))$	$1/4(p(B_i)p(AB_m) + p(AB_i)p(B_m))$	0
BB-AB	$(-a - d + \xi_j)^2$	$1/2(p(AB_i)p(BB_m) + p(BB_i)p(AB_m))$	$1/4(p(B_i)p(AB_m) + p(AB_i)p(B_m))$	0
AA-BB	$(2a + \xi_j)^2$	$1/4 p(AB_i) p(AB_m)$	0	0
BB-AA	$(-2a + \xi_j)^2$	$1/4 p(AB_i) p(AB_m)$	0	0
Total		1	1	1

Thus, we deduced the formula for covariance between hybrid sibs as the weighted sum of three (additive, dominance and residual) components:

$$\text{Cov}(X_{1j}, X_{2j} | \hat{\pi}_{Mj}) = \mathbf{f}(\hat{\pi}_{Mj}, \theta) \mathbf{A} + \mathbf{g}(\hat{\pi}_{Mj}, \theta) \mathbf{D} + \mathbf{R}. \quad (9)$$

One can conclude that trait covariance depends on the necessary set of parameters $\{a, d, p_1, p_2, \theta\}$.

Criterion for the definition of QTL position

To localize a QTL on a chromosome the maximum likelihood method was used. This method enables to choose the most suitable genetic model, estimate the modelling parameters and define the position of the QTL with the required accuracy. Note that if there are no genetic effects ($a = 0$ and $d = 0$), it is impossible to localize a QTL since the recombination fractions between the QTL and markers can not be estimated. Let us consider two genotyped markers flanking the QTL and construct the suitable log-likelihood function:

$$\ln L = \text{const} - 1/2 \sum [\ln |\mathbf{V}| + (\mathbf{X} - \mathbf{E}_X) \mathbf{V}^{-1} (\mathbf{X} - \mathbf{E}_X)^T],$$

where the summation is over the two flanking markers; \mathbf{X} and \mathbf{E}_X are horizontal vectors of quantitative trait values and their expectations, respectively; \mathbf{V} is a covariance matrix with the elements $\text{Cov}(X_{1j}, X_{2j} | \hat{\pi}_{Mj})$. The log-likelihood function does not change a form at multiple analyses, because to localize the QTL among multiple markers, it is necessary to test each chromosome fragment bracketed by only two adjacent genotyped markers.

We constructed the statistics as a double likelihood ratio, $2(\ln L_1 - \ln L_0)$, where L_0 is the maximum likelihood under a null hypothesis H_0 , obtained by imposing restrictions on certain parameters of interest, and L_1 is the maximum likelihood under an alternative hypothesis H_1 , where these restrictions are removed. Here, we have chosen hypothesis H_1 in which the parameter θ is not fixed, and hypothesis H_0 in which the parameter θ is equal to the fixed value, θ_k . One can let the recombination frequency between one of the markers and the QTL, θ_k , be correlated with genetic distance, k , by the Kocambi mapping function [27], and take into account interference:

$$\theta_k = 1/2 (e^{4k} - 1)/(e^{4k} + 1),$$

where k varies from 0 to r discretely (with given step length), and r is a fixed genetic distance between two markers. Thus, we have several null hypotheses from which it is necessary to choose a suitable one. If the value of the statistics is calculated for each probable k -position

of the QTL and compared with the critical value, then we can accept or reject the given position as correct. Indeed, the specified criterion is the linkage test, for which the critical value transformed from LOD score is equal to $2\ln(10^3) = 13.8$. Note that in spite of the fact that many authors have demonstrated that, for evidence of more significant linkage, LOD score threshold is greater than 3, we use just this traditional threshold as being more convenient for comparison of our method with other ones with same LOD score thresholds. But researchers can choose a more severe threshold.

Power

From mathematical statistics it is known that the likelihood ratio test has a central χ^2 -distribution under a null hypothesis and a noncentral χ^2 -distribution under an alternative hypothesis in large samples [25]. Given a critical P -value, the power of a χ^2 -test can be determined from the noncentrality parameter, λ , which is directly proportional to the sample size, N , and to the degree of freedom of the noncentral χ^2 -distribution, df . To estimate the power for any sample size at a given λ and df , one can refer to the appropriate function of the noncentral χ^2 -distribution. It is possible to derive analytical formulas for the noncentrality parameter without carrying out data simulation [28]. For this, it is necessary to obtain the asymptotical values of the maximum-likelihood estimates of parameters under both the H_0 and H_1 hypotheses, and then to take the log-likelihood expectations under these hypotheses evaluated at their respective asymptotical parameter estimations. The noncentrality parameter is then:

$$\lambda = E(2\ln L_1) - E(2\ln L_0). \quad (10)$$

The linkage test is caused by distinctions only in covariance matrixes, \mathbf{V} , according to the marker IBD-distribution. For example, we constructed a noncentrality parameter for sibs since they have identical variance components. For notational convenience, we assume that the quantitative trait has unit variance, so that V_A , V_D , V_R and V_e represent both the variances and the proportions of variance. Then under the H_0 hypothesis, when k varies from 0 to r , the asymptotic estimations of covariance are:

$$V_0(i, j) = \begin{cases} V_A + V_D + V_R + V_e, & \text{if } i = j, \\ \mathbf{f}(\hat{\pi}_M, \theta = \theta_k) V_A + \mathbf{g}(\hat{\pi}_M, \theta = \theta_k) V_D + V_R, & \text{if } i \neq j, \end{cases}$$

and under the alternative H_1 hypothesis, when θ is not fixed, they are:

$$V_1(i, j) = \begin{cases} V_A + V_D + V_R + V_e, & \text{if } i = j, \\ \mathbf{f}(\hat{\pi}_M, \theta) V_A + \mathbf{g}(\hat{\pi}_M, \theta) V_D + V_R, & \text{if } i \neq j, \end{cases}$$

In the presence of marker information, $E(2\ln L_k)$, from formula (10) for both hypotheses ($k = 0$ or 1) is calculated as [28]:

$$E(2\ln L_k) = -E(\ln |V_k|) - E(X - E_X)V_k^{-1}(X - E_X)^T \\ = -\sum_i p_i \ln |V_{ki}| - s,$$

where s is sibship size; and p_i and V_{ki} are the probability and the covariance matrix for the i th marker genotype configuration, respectively.

We defined the noncentrality parameter for the sib-pair and then generalized it for an entire sibship. For any relative pair, one can unambiguously assign three covariance matrices z_π differing by a portion of IBD-alleles on the QTL, $\pi = 0, 1/2$ or 1 . From formula (9) it follows that their non-diagonal elements are equal to $z_{\pi=0}(1,2) = V_R$, $z_{\pi=0.5}(1,2) = 1/2 V_A + V_R$ and $z_{\pi=1}(1,2) = V_A + V_D + V_R$, and diagonal elements are obviously equal to 1. In a random sample of sib-pairs, these covariance matrices are expected to occur in the proportions $1/4:1/2:1/4$, so that the noncentrality parameter for sibship is:

$$\lambda = E(\ln |V_0|) - E(\ln |V_1|) \\ = 1/4 \ln |V_{0,\pi=0}| + 1/2 \ln |V_{0,\pi=0.5}| + 1/4 \ln |V_{0,\pi=1}| - 1/4 \ln |V_{1,\pi=0}| - 1/2 \ln |V_{1,\pi=0.5}| - 1/4 \ln |V_{1,\pi=1}|.$$

The conditional sib-pair correlations of trait values, given the IBD-status at the QTL, can be deduced from the conditional distribution $\Pr(\pi_{QTLj} | \pi_{Mij})$ as:

$$c_0 = V_R + (1 - \Psi)V_A + (1 - \Psi)^2 V_D, \\ c_1 = V_R + 1/2 V_A + \Psi(1 - \Psi)V_D, \\ c_2 = V_R + \Psi V_A + \Psi^2 V_D, \quad (11)$$

where $\Psi = \theta^2 + (1 - \theta)^2$. The noncentrality parameter for sib-pair is then given by:

$$\lambda = 1/4 \ln(1 - c_0^2) + 1/2 \ln(1 - c_1^2) + 1/4 \ln(1 - c_2^2) - 1/4 \ln(1 - c_0'^2) - 1/2 \ln(1 - c_1'^2) - 1/4 \ln(1 - c_2'^2),$$

where c'_i are values of c_i ($i = 0, 1$, and 2) at the fixed Ψ_k , $\Psi_k = \theta_k^2 + (1 - \theta_k)^2$.

When the first-order approximation, $\ln(1-x) \approx -x$, is used for small values of x , the formula for the noncentrality parameter becomes the simplest:

$$\lambda_1 \approx 1/4 (c_0'^2 - c_0^2) + 1/2 (c_1'^2 - c_1^2) + 1/4 (c_2'^2 - c_2^2) \\ = \Delta_1 \Delta_2 \{ 1/2 V_A^2 + V_D V_A + 1/2 (1 + \Delta_1^2 + \Delta_2^2) V_D^2 \},$$

where $\Delta_1 = \Psi - \Psi_k$ and $\Delta_2 = 1 - \Psi_k - \Psi$. In this case λ_1 is proportional only to the squares and products of the additive V_A and dominance V_D variance components and does not

depend on the component conditioned by the hybrid (inter-population) origin of the sibs.

For a sufficiently accurate calculation of the noncentrality parameter, as often as not the second-order approximation, $\ln(1-x) \approx -x - 1/2 x^2$, is used. It follows therefore that

$$\lambda \approx \lambda_1 + 1/8 (c_0'^4 - c_0^4) + 1/4 (c_1'^4 - c_1^4) + 1/8 (c_2'^4 - c_2^4). \quad (12)$$

The analytical expression for the noncentrality parameter after substituting the expressions (11) in formula (12) is lengthy, but we can see that λ depends on all variance components, V_A , V_D , and V_R . Moreover, the power to detect a given QTL effect increases with increasing proportion of the residual component, V_R . To obtain more accurate results, it is possible to use an approximation by involving higher-order terms.

To determine a noncentrality parameter for the entire sibship, we used a suitable approximation to calculate a determinant of correlation matrix (non-singular and symmetric) as shown in [28]:

$$\ln |V| \approx \ln(1 - \sum V_{jk}^2) = -\sum V_{jk}^2, \quad (13)$$

where \sum denotes the sum over all possible sib-pairs (j, k), $j < k$. Then for an s -size sibship the noncentrality parameter, λ_s , is equal to:

$$\lambda_s \approx 1/2 s(s-1) \lambda, \quad (14)$$

where $1/2s(s-1)$ is the number of sib-pairs. As is obvious, the noncentrality parameter for the linkage test is proportional to the number of all pairs in the sibship. It is noteworthy that formula (14) is not exact for small samples, and in this case, it is necessary to calculate the power through data simulation.

In the case of analyses of a hybrid pedigree of arbitrary structure, the noncentrality parameter can be obtained in a similar manner. For this purpose, noncentrality parameters are calculated for all relative pairs of the pedigree analysed, and are then summarized according to approximation (13). When the theoretical noncentrality parameter has been obtained, it is easy to calculate the size of the sample required for any required level of significance and power. For the linkage test, the level of significance required is traditionally set at a LOD score of 3, which is equivalent to a χ^2 statistics of 13.8 with $df = 2$ and a fixed-sample one-tailed significance level of 0.0001. The noncentrality parameter required for 80% power is 20.8 [28]. For example, the number of sib-pairs required can be obtained by dividing the noncentrality parameter

Table 6: Sample sizes required for 80% power to detect linkage for the range of V_R , recombination fraction and fixed components $V_D = 0.15$ and $V_A = 0.15$

Recombination fraction, θ	Sample size required for $V_R =$							
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35
0.00	1530	1460	1379	1291	1200	1110	1023	940
0.05	2453	2346	2220	2082	1938	1793	1653	1520
0.10	4078	3909	3704	3478	3240	3001	2767	2545
0.15	7142	6855	6504	6111	5697	5279	4871	4480

required (i.e. 20.8 for the linkage test) by the theoretical noncentrality parameter per sib-pair.

Table 6 demonstrates the sample sizes required for 80% power at the critical value of 13.8 for a range of variance components, V_R , and Ψ . Additive and dominance components are assumed to be 0.15. Sham et al. [28] have shown that detection of QTL by linkage is only feasible if the proportion of QTL variance considered is 10% or more. At this level of QTL variance, more than 20000 sib-pairs are required for linkage analysis.

Simulations

To examine the performance of the proposed approach in realistic situations we conducted simulation studies on examples of inbred sibships. We generated 10, 20, 30, and 40 hybrid pedigrees covering three generations of individuals (F_0 , F_1 , and F_2 generations). Founders (F_0 individuals) of each pedigree analysed are two individuals from the P_1 population and single individual from the P_2 population. The founders from different populations formed a crossing pair and had one offspring for the F_1 generation. Inbred crossing between related F_1 individuals contributed to the F_2 generation by the size of 10 offsprings. Thus, each pedigree consisted of three founders, two F_1 individuals, and ten F_2 individuals. For all inbred sibs from the F_2 generation, the marker genotypes and phenotypic values of the trait were simulated and were considered as known.

We considered two QTL positions between two markers, in turn assuming that the QTL is located in the chromosome positions 5 and 10 cM, and that the markers flanking it are fixed at positions 0 and 25 cM. We specified the distribution of allele frequencies of the QTL for all the individuals analysed based on the assumption that the Hardy-Weinberg equilibrium is carried out for the founders.

Let there be four-allelic markers (two unique alleles from each initial population). The marker genotypes of the founders were selected randomly, assuming an even dis-

tribution of frequencies of marker alleles. The additive and dominance genetic values were taken to be $a = 3$ and $d = 1$, respectively. For the founders, the QTL allelic frequencies were taken to be $p_1 = 0.9$ and $p_2 = 0.4$. The phenotypic values were obtained by adding the normal deviation $N(0,1)$ to the genetic value. For each given set of model parameters $\{p_1, p_2, a, d, \theta\}$ and the given sample size, 100 replicates were simulated.

Our purpose is to locate the QTL estimating allelic frequencies for initial populations P_1 and P_2 and the genetic effects of the trait in question in each replicate. To locate the QTL on a chromosome fragment between markers, we discretely moved along the fragment at a step length of 0.01 cM and estimated the double likelihood ratio statistics at each point. If the statistics calculated at a point was higher than the critical value, then the hypothesis of the localization of the QTL at this point was rejected. The QTL was hypothetically located at the point where the statistics had the lowest value.

We compared our method with the method for QTL analysis of F_2 crosses between outbred lines, described in [2,29], which was performed using the Qxpack software available free at [30].

The performance of both methods was tested using the same simulated data. The comparative characteristics were the frequency of the events consisting in the fact that the true location of the QTL would not be rejected (W_1), and the frequency of the events consisting in the fact that the statistical test would indicate the true QTL location as the most likely one (W_2). It is obvious that $W_1 \geq W_2$. It should be noted that the value of $(1-W_1)$ can be interpreted as type I error rate, and value of W_2 can be analogous to power of the method.

We compared the W_1 and W_2 frequencies for the two methods at different sample sizes, $N_{\text{ped}} = 150, 300, 450$, and 600. Figures 2 and 3 show the results of simulation studies for two positions of the QTL, 5 and 10 cM, respec-

tively. As can be seen, at both positions of the QTL, the statistical characteristics W_1 and W_2 calculated by the method proposed are higher than those calculated using Qxpak at all sample sizes considered.

When the QTL is really localized at the position 5 cM (Figure 2), our method demonstrates the highest values of W_1 frequency, exceeding 0.8 at any sample size. Only beginning at $N_{\text{ped}} = 450$, each of bar graphs for W_1 gets in a range of 5-percentage error of another. Regarding the W_2 characteristic, its values do not exceed 0.8 for either method at any sample size. When the QTL is localized at position 10 cM (Figure 3), the frequencies W_1 and W_2 are higher than 0.95 and 0.8, respectively, beginning at $N_{\text{ped}} = 150$ for our method and beginning at $N_{\text{ped}} > 600$ for Qxpak.

All aforesaid facts speak in favour for our method for the analysis of hybrid pedigrees with dominance and inbreeding effects. The results obtained justify the QTL analysis by our method that yields more accurate data on the localization of the QTL.

Discussion

In this study we have updated the variance-components method for the analysis of hybrid pedigrees with dominance and inbreeding. We have considered hybrid sibships as an example to demonstrate the method. An advantage of our method is to partition into variance

components, where one of the components is conditioned by the inter-population origin of individuals and inbreeding. There is no necessity to resolve this component into separate elements caused by inter-population origin and separate elements caused by inbreeding since these elements are indivisible in variances and covariances and therefore can not be estimated singly.

We have derived an intuitively appealing result regarding the power of our method under a variance-components model for larger samples of sibships. If the effects of QTL are small, the results are particularly simple. We have generally arrived at the conclusion that the power of our method decreases rapidly with decreasing proportion of the variance component caused by the hybrid origin and by inbreeding. This means that the sample size required for 80% power for hybrid pedigrees is less than for pedigrees descended from one pure line.

For simplicity, we did not consider such fixed factors affecting the formation of traits as for example, age and sex, but these factors can easily be included in our model. Moreover, the method suggested can be used to choose the most suitable model for the description of the data: additive models ($d = 0$), dominance models ($a = 0$), models of crosses of two pure lines ($p_1 = 0$, and $p_2 = 1$) or models of intra-population crosses ($p_1 = p_2$).

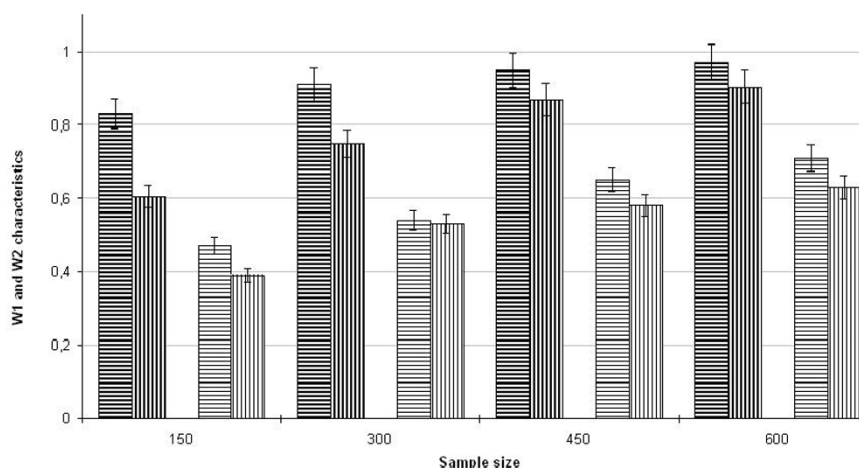


Figure 2

Bar graphs for W_1 and W_2 statistical characteristics depending on the sample size at the QTL position 5 cM.

Bar graphs hatched by bold lines correspond to characteristic W_1 ; bar graphs hatched by thin lines correspond to characteristic W_2 ; vertical lines are intended for the method realized by software Qxpak; horizontal lines are intended for the method proposed in this study.

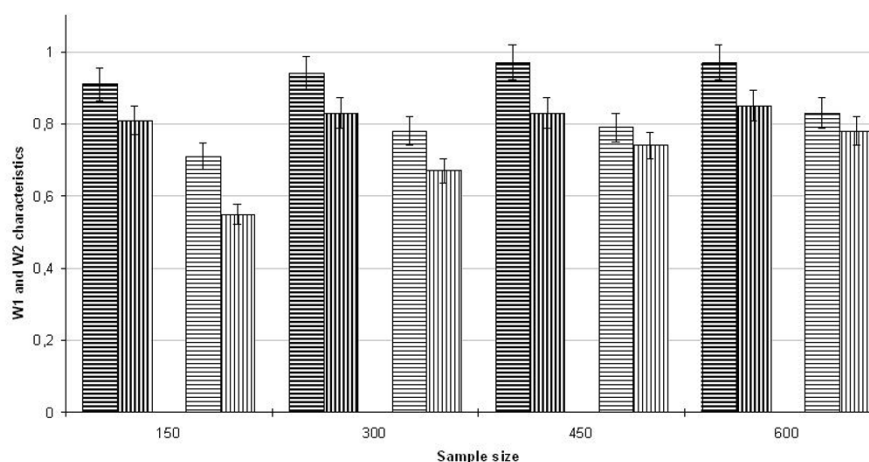


Figure 3

Bar graphs for W_1 and W_2 statistical characteristics depending on the sample size at the QTL position 10 cM. Bar graphs hatched by bold lines correspond to characteristic W_1 ; bar graphs hatched by thin lines correspond to characteristic W_2 ; vertical lines are intended for the method realized by software Qxpak; horizontal lines are intended for the method proposed in this study.

The results obtained make it possible to draw conclusions on the competence of the incorporated analysis that could specify not only the localization of the QTL, but also an estimate of the values of QTL effects.

Conclusion

We have presented a new modification of the variance-components method for QTL analysis. It is a linkage test method, whose originality consists in considering the trait effect caused by inter-population origin and inbreeding. Analytical derivations for the variance components make it possible to analyse their dependence from the model parameters.

The analytical expressions for the power of our method avoid the intensive computations required for simulated data processing and allow to estimate the size of the pedigree required. We have shown that the method is more powerful if the QTL effects conditioned by inter-population origin and inbreeding are increased. Several improvements can be developed to take into account fixed factors affecting trait formation, such as age and sex.

Our method uses the trait values and the marker information for each individual of a pedigree with an arbitrary

structure including inbred loops as initial data and can be valuable for fine mapping purposes.

Methods

To choose the most correct genetic model for quantitative trait formation, estimate the modelling parameters, and define the position of the QTL on a chromosome with the accuracy required, we employed statistics based on likelihood maximization:

$$\max_{a,d,p_1,p_2} \ln L(a,d,p_1,p_2,\theta | \hat{\pi}_{Mj})$$

This maximization is numerically carried out using the simplex METHI – program specifically developed to obtain maximum likelihood (ML) and ML-parameter estimates of likelihood function. METHI uses a method of configurations when maximising a function. We have free access to METHGI on our laboratory website [31]. The parameters $\{a, d, p_1, p_2\}$ must be estimated. For the recombination frequency, we have assigned different fixed values corresponding to specific distances on the chromosome.

Appendix

Let the relationship degree between the father (i_f) and the mother (i_m) of an inbred individual (i) be equal to $(1/2)^k$, and the inbreeding parameter, τ_k , be calculated by formula (3). Using the mathematical induction method, we show that $\tau_{k+1} = 1/2\tau_k$, if a single non-inbred ($k+1$)-relative is to be inserted into the inbred loop. For certainty, we added the new member between individuals i and i_f . Consequently, we re-denoted the i th individual as the j th individual because his genotype distribution had changed. Apparently, the renamed j th individual has the former mother $j_m = i_m$, and his father is the added relative denoted as j_f . In one turn, the father of the new member is the father of the i th individual $j_{ff} = i_f$, and the mother of the new member is the individual j_{fm} , unrelated to the father. It is easy to find the frequency of genotype AA of the offspring j analysed at the outbred cross:

$$\begin{aligned} p(AA_j) &= p(A_{jm})p(A_{jf}) \\ &= p(A_{jm})(p(A_{if}) + p(A_{ifm}))/2, \end{aligned}$$

and at the inbred cross:

$$\begin{aligned} p^{\text{inb}}(AA_j) &= \sum_{G_o} p(AA_j | G_o)p(G_o) \\ &= \sum_{G_o} p(A_{im} | G_o)p(A_{jf} | G_o)p(G_o), \end{aligned}$$

where G_o designates the set of genotypes of the common ancestors for parents of the j th individual. We expressed the frequencies of genotypes of the father j_f through frequencies of his parents:

$$\begin{aligned} p^{\text{inb}}(AA_j) &= \frac{1}{2} \sum_{G_o} p(A_{im} | G_o)[p(A_{jff} | G_o) + p(A_{jfm})]p(G_o) \\ &= \frac{1}{2} \sum_{G_o} p(A_{im} | G_o)[p(A_{if} | G_o) + p(A_{jfm})]p(G_o). \end{aligned}$$

After of some transformations, it is clear that the frequency of genotype AA at the inbred cross can be expressed through a similar frequency of the individual i th:

$$p^{\text{inb}}(AA_j) = 1/2(p^{\text{inb}}(AA_i) + p(A_{jfm})p(A_{im})).$$

The inbreeding parameter τ_{k+1} is then:

$$\begin{aligned} \tau_{k+1} &= p^{\text{inb}}(AA_j) - p(AA_i) \\ &= \frac{1}{2}(p^{\text{inb}}(AA_i) + p(A_{jfm})p(A_{im}) - p(A_{jm})[p(A_{if}) + p(A_{ifm})]) \\ &= \frac{1}{2}(p^{\text{inb}}(AA_i) - p(AA_i)) \\ &= \frac{1}{2}\tau_k. \end{aligned}$$

Thus, we have shown that the inbreeding parameter for the individual descended from any type of inbred cross, depends on the degree of relationship of his parents and

genotype distribution of the common ancestors, and does not depend on the distribution of the inbred offspring.

As a result, inbreeding changes the parameters of distribution of quantitative trait values for hybrid individuals: genotypic means decrease or are constant, and covariances basically increase. It is important to note that the account of inbreeding of hybrid individuals does not complicate QTL analysis, and more exactly estimates parameters of distribution of quantitative trait values.

Acknowledgements

I gratefully thank Anne-Lise Haenni for critical reading of this manuscript, Tatiana Axenovich for the helpful discussions, and Galina Karpova, Dmitry Graifer and Ivan Shatsky for their help in the preparation of this manuscript.

References

- Haley CS, Knott SA, Elsen JM: **Mapping quantitative trait loci in crosses between outbred lines using least squares.** *Genetics* 1994, **136**:1195-1207.
- Pérez-Enciso M, Varona L: **Quantitative trait loci mapping in F (2) crosses between outbred lines.** *Genetics* 2000, **155**:391-405.
- Pérez-Enciso M, Fernando R, Bidanel J-P, Le Roy P: **Quantitative trait locus analysis in crosses between outbred lines with dominance and inbreeding.** *Genetics* 2001, **159**:413-422.
- Pérez-Enciso M, Clop A, Folch JM, Sanchez A, Oliver MA, Ovilo C, Barragan C, Varona L, Noguera JL: **Exploring alternative models for sex-linked quantitative trait loci in outbred populations: application to an Iberian × landrace pig intercross.** *Genetics* 2002, **161**:1625-1632.
- Ovilo C, Clop A, Noguera JL, Oliver MA, Barragán C, Rodríguez C, Silió L, Toro MA, Coll A, Folch JM, Sánchez A, Babot D, Varona L, Pérez-Enciso M: **Quantitative trait locus mapping for meat quality traits in an Iberian × Landrace F2 pig population.** *J Anim Sci* 2002, **80**:2801-2808.
- Chevalet C, Gillois M, Khang VT: **Conditional probabilities of identity of genes at a locus linked to a marker.** *Genetic Selection Evolution* 1984, **16**:431-444.
- Amos CI, Elston RC: **Robust methods for the detection of genetic linkage for quantitative data from pedigrees.** *Genet Epidemiol* 1989, **6**:349-360.
- Goldgar DE: **Multipoint analysis of human quantitative genetic variation.** *Am J Hum Genet* 1990, **47**:957-967.
- Schork NJ: **Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations.** *Am J Hum Genet* 1993, **53**:1306-1319.
- Amos CI: **Robust variance-components approach for assessing in pedigree.** *Am J Hum Genet* 1994, **54**:535-543.
- Blangero J, Almasy L: **Multipoint oligogenic linkage analysis of quantitative traits.** *Genet Epidemiol* 1997, **14**:959-964.
- Williams JT, Duggirala R, Blangero J: **Statistical properties of a variance-components method for quantitative trait linkage analysis in nuclear families and extended pedigrees.** *Genet Epidemiol* 1997, **14**:1065-1070.
- Blangero J, Williams JT, Almasy L: **Robust LOD scores for variance component-based linkage analysis.** *Genet Epidemiol* 2000:8-12.
- Lange K, Westlake J, Spence MA: **Extensions to pedigree analysis III. Variance components by the scoring method.** *Ann Hum Genet* 1976, **39**:485-491.
- Almasy L, Blangero J: **Multipoint quantitative-trait linkage analysis in general pedigrees.** *Am J Hum Genet* 1998, **62**:1198-1211.
- Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J: **A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2.** *Nat Genet* 1997, **15**:273-276.
- Lander ES, Botstein D: **Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps.** *Genetics* 1989, **121**:185-199.

18. Haley CS, Knott SA: **A simple regression method for mapping quantitative trait loci in line crosses using flanking markers.** *Heredity* 1992, **69**:315-324.
19. Martinez O, Curnow RN: **Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers.** *Theor Appl Genet* 1992, **85**:480-488.
20. Jansen R: **Interval mapping of multiple quantitative trait loci.** *Genetics* 1993, **135**:205-211.
21. Xie C, Gessler DD, Xu S: **Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method.** *Genetics* 1998, **149**:1139-1146.
22. Lo LL, Fernando RL, Cantet RJC, Grossman M: **Theory for modeling means and covariances in a two-breed population with dominance inheritance.** *Theor Appl Genet* 1995, **90**:49-62.
23. Axenovich TI: **Inheritance of Quantitative Traits in Hybrid Pedigrees: Mixed Models.** *Russian Journal of Genetics* 1999, **35**(4):530-539.
24. Li CC: *First course in population genetics* Pacific grove (California): The boxwood press; 1976.
25. Kendall M, Stuart A: *The advanced theory of statistics Volume 2. Inference and relationship.* New York: John Wiley and Sons; 1979.
26. Haseman JK, Elston RC: **The investigation of linkage between a quantitative trait and a marker locus.** *Behav Genet* 1972, **2**:3-19.
27. Kosambi DD: **The estimation of map distances from recombination values.** *Ann Eugen* 1922, **12**:172-175.
28. Sham PC, Cherny SS, Purcell S, Hewitt JK: **Power of linkage versus association analysis of quantitative traits, by use variance-components models, for sibship data.** *Am J Hum Genet* 2000, **66**:1616-1630.
29. Perez-Enciso M, Misztal I: **Qxpak: a versatile mixed model application for genetical genomics and QTL analyses.** *Bioinformatics* 2004, **20**:2792-2798.
30. [<http://www.icrea.es/pag.asp?id=Miguel.Perez>].
31. [<http://mga.bionet.nsc.ru/soft/methgi/methgi.html>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

